Variations of the Morse-Hedlund Theorem for k-Abelian Equivalence

Juhani Karhumäki[1][*], Aleksi Saarela[1][*], and Luca Q. Zamboni[12][**]

[1] Department of Mathematics and Statistics & FUNDIM, University of Turku,
FI-20014 Turku, Finland
[2] Université de Lyon, Université Lyon 1, CNRS UMR 5208, Institut Camille Jordan,
43 boulevard du 11 novembre 1918, F69622 Villeurbanne Cedex, France
[3] `karhumak@utu.fi`, `amsaar@utu.fi`, `lupastis@gmail.com`

Abstract. In this paper we investigate local to global phenomena for a new family of complexity functions of infinite words indexed by $k \in \mathbb{N}_1 \cup \{+\infty\}$ where $\mathbb{N}_1$ denotes the set of positive integers. Two finite words u and v in A^* are said to be k-Abelian equivalent if for all $x \in A^*$ of length less than or equal to k, the number of occurrences of x in u is equal to the number of occurrences of x in v. This defines a family of equivalence relations $\sim_k$ on A^*, bridging the gap between the usual notion of Abelian equivalence (when $k = 1$) and equality (when $k = +\infty$). Given an infinite word $w \in A^\omega$, we consider the associated complexity function $\mathcal{P}_w^{(k)} : \mathbb{N}_1 \to \mathbb{N}_1$ which counts the number of k-Abelian equivalence classes of factors of w of length n. As a whole, these complexity functions have a number of common features: Each gives a characterization of periodicity in the context of bi-infinite words, and each can be used to characterize Sturmian words in the framework of aperiodic one-sided infinite words. Nevertheless, they also exhibit a number of striking differences, the study of which is one of the main topics of our paper.

1 Introduction

A fundamental problem in both mathematics and computer science is to describe local constraints which imply global regularities. A splendid example of this phenomena may be found in the framework of combinatorics on words. In their seminal papers [13,14], G. A. Hedlund and M. Morse proved that a bi-infinite word w is periodic if and only if for some positive integer n, w contains at most n distinct factors of length n. In other words, it describes the exact borderline between periodicity and aperiodicity of words in terms of the *factor complexity function* which counts the number of distinct factors of each length n. An analogous result was established some thirty years later by E. Coven and G.A. Hedlund in the framework of Abelian equivalence. They show that a bi-infinite word is periodic if and only if for some positive integer n all factors of

---

w are Abelian equivalent. Thus once again it is possible to distinguish between periodic and aperiodic words on a local level by counting the number of Abelian equivalence classes of factors of length n.

In this paper we study the local to global behavior for a new family of complexity functions $\mathcal{P}_w^{(k)}$ of infinite words indexed by $k \in \mathbb{N}_1 \cup \{+\infty\}$ where $\mathbb{N}_1 = \{1, 2, 3, \ldots\}$ denotes the set of positive integers. Let $k \in \mathbb{N}_1 \cup \{+\infty\}$ and A be a finite non-empty set. Two finite words u and v in A^* are said to be *k-Abelian equivalent* if for all $x \in A^*$ of length less than or equal to k, the number of occurrences of x in u is equal to the number of occurrences of x in v. This defines a family of equivalence relations $\sim_k$ on A^*, bridging the gap between the usual notion of Abelian equivalence (when $k = 1$) and equality (when $k = +\infty$). Abelian equivalence of words has long been a subject of great interest (see for instance Erdös problem, [1,2,3,5,11,15,16,17,18]). Although the notion of k-Abelian equivalence is quite new, there are already a number of recent papers on the topic [8,6,7,9,12,10].

Given an infinite word $w \in A^\omega$, we consider the associated complexity function $\mathcal{P}_w^{(k)} : \mathbb{N}_1 \to \mathbb{N}_1$ which counts the number of k-Abelian equivalence classes of factors of w of length n. Thus $\mathcal{P}_w^{(\infty)}$ corresponds to the usual factor complexity while $\mathcal{P}_w^{(1)}$ corresponds to Abelian complexity. As it turns out, each intermediary complexity function $\mathcal{P}_w^{(k)}$ for $2 \le k < +\infty$ can be used to detect periodicity of words. In order to describe this connection, we will make use of the following auxiliary function first discovered in [10]:

$$q^{(k)}(n) = \begin{cases} n+1 & \text{if } n \le 2k - 1 \\ 2k & \text{if } n \ge 2k \end{cases}.$$

As a starting point of our research, we list two classical results on factor and Abelian complexity in connection with periodicity, and their k-Abelian counterparts proved by the authors in [10]. We note that in each case, the first two items are included in the third.

Theorem 1. *Let w be a bi-infinite word over a finite alphabet. Then the following properties hold:*

- *(M. Morse, G.A. Hedlund, [13]) The word w is periodic if and only if $\mathcal{P}_w^{(\infty)}(n) \le n$ for some $n \ge 1$.*
- *(E.M. Coven, G.A. Hedlund, [2]) The word w is periodic if and only if $\mathcal{P}_w^{(1)}(n) = 1$ for some $n \ge 1$.*
- *The word w is periodic if and only if $\mathcal{P}_w^{(k)}(n) < q^{(k)}(n)$ for some $k \in \mathbb{N}_1 \cup \{+\infty\}$ and $n \ge 1$.*

Also, each complexity provides a characterization for an important class of binary words, the so-called *Sturmian* words:

Theorem 2. *Let w be an aperiodic one-sided infinite word. Then the following properties hold:*

- (M. Morse, G.A. Hedlund, [14]). The word w is Sturmian if and only if $\mathcal{P}_w^{(\infty)}(n) = n + 1$ for all $n \geq 1$.
- (E.M. Coven, G.A. Hedlund, [2]). The word w is Sturmian if and only if $\mathcal{P}_w^{(1)}(n) = 2$ for all $n \geq 1$.
- The word w is Sturmian if and only if $\mathcal{P}_w^{(k)}(n) = q^{(k)}(n)$ for all $k \in \mathbb{N}_1 \cup \{+\infty\}$ and $n \geq 1$.

However, in other respects, these various complexities exhibit radically different behaviors. For instance, in the context of one-sided infinite words, the first item in Theorem 1 gives rise to a characterization of ultimately periodic words, while for the other two, the result holds in only one direction: If $\mathcal{P}_w^{(k)}(n) < q^{(k)}(n)$ for some $k \in \mathbb{N}_1$ and $n \geq 1$ then w is ultimately periodic, but not conversely (see [10]). For instance in the simplest case when $k = 1$, it is easy to see that if w is the ultimately periodic word 01^ω, then for each positive integer n there are precisely two Abelian classes of factors of w of length n. However, the same is true of the (aperiodic) *Fibonacci infinite word*

$$w = 010010100100101001\ldots$$

defined as the fixed point of the morphism $0 \mapsto 01$, $1 \mapsto 0$. Analogously, in Theorem 2 the first item holds true without the added assumption that w be aperiodic, while the other two items do not. Another striking difference between them is in their rate of growth. Consider for instance the binary Champernowne word

$$\mathcal{C} = 011011100101110111110001001\ldots$$

obtained by concatenating the binary representation of the consecutive natural numbers. Let w denote the morphic image of $\mathcal{C}$ under the Thue-Morse morphism τ defined by $0 \mapsto 01$ and $1 \mapsto 10$. Then while $\mathcal{P}_w^{(\infty)}(n)$ has exponential growth, it can be shown that $\mathcal{P}_w^{(1)}(n) \leq 3$ for all n. Yet another fundamental disparity concerns the difference $\mathcal{P}_w^{(k)}(n + 1) - \mathcal{P}_w^{(k)}(n)$. For the factor complexity, one always has $\mathcal{P}_w^{(\infty)}(n + 1) - \mathcal{P}_w^{(\infty)}(n) \geq 0$, while for general k this inequality is far from being true.

A primary objective in this paper is to study the asymptotic lower and upper complexities defined by

$$\mathcal{L}_w^{(k)}(n) = \min_{m \geq n} \mathcal{P}_w^{(k)}(m) \qquad \text{and} \qquad \mathcal{U}_w^{(k)}(n) = \max_{m \leq n} \mathcal{P}_w^{(k)}(m).$$

Surprisingly these quantities might deviate from one another quite drastically. Indeed, one of our main results is to compute these values for the famous Thue-Morse word. We show that the upper limit is logarithmic, while the lower limit is just constant, in fact at most 8 in the case $k = 2$. This is quite unexpected considering the Thue-Morse word is both pure morphic and Abelian periodic (of period 2). If we however allow more general words, then we obtain much stronger evidence of the non-existence of gaps in low k-Abelian complexity classes. We construct uniformly recurrent infinite words having arbitrarily low

upper limit and just constant lower limit. The concept of k-Abelian complexity also leads to many interesting open questions. We conclude the paper in Sect. 6 by mentioning some of these problems.

2   Preliminaries

Let Σ be a finite non-empty set called the *alphabet*. The set of all finite words over Σ is denoted by Σ^* and the set of all (right) infinite words is denoted by Σ^ω. The set of positive integers is denoted by $\mathbb{N}_1$.

Let $w \in \Sigma^\omega$. The word w is *periodic* if there is $u \in \Sigma^*$ such that $w = u^\omega$, and *ultimately periodic* if there are $u, v \in \Sigma^*$ such that $w = vu^\omega$. If w is not ultimately periodic, then it is *aperiodic*. Let $u = a_0 \ldots a_{m-1}$ and $a_0, \ldots, a_{m-1} \in \Sigma$. The *prefix* of length n of u is $\mathrm{pref}_n(u) = a_0 \ldots a_{n-1}$ and the *suffix* of length n of u is $\mathrm{suff}_n(u) = a_{m-n} \ldots a_{m-1}$. If $0 \leq i \leq m$, then the notation $\mathrm{rfact}_n^i(u) = a_i \ldots a_{i+n-1}$ is used. The length of a word u is denoted by $|u|$ and the number of occurrences of another word x as a factor of u by $|u|_x$. Two words $u, v \in \Sigma^*$ are *Abelian equivalent* if $|u|_a = |v|_a$ for all $a \in \Sigma$.

Let $k \in \mathbb{N}_1$. Two words $u, v \in \Sigma^*$ are *k-Abelian equivalent* if $|u|_x = |v|_x$ for all words x of length at most k. k-Abelian equivalence is denoted by $\sim_k$. If the length of u and v is at least $k-1$, then $u \sim_k v$ if and only if $|u|_x = |v|_x$ for all words x of length k and $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$ and $\mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$.

Let $w \in \Sigma^\omega$. The set of factors of w of length n is denoted by $\mathcal{F}_w(n)$. The *factor complexity* of w is the function $\mathcal{P}_w^{(\infty)} : \mathbb{N}_1 \to \mathbb{N}_1$ defined by

$$\mathcal{P}_w^{(\infty)}(n) = \#\mathcal{F}_w(n).$$

Let $k \in \mathbb{N}_1$. The *k-Abelian complexity* of w is the function $\mathcal{P}_w^{(k)} : \mathbb{N}_1 \to \mathbb{N}_1$ defined by

$$\mathcal{P}_w^{(k)}(n) = \#(\mathcal{F}_w(n)/\sim_k).$$

Factor complexity functions are always increasing, and even strictly increasing for aperiodic words. For k-Abelian complexity this is not true. This is why we define *upper k-Abelian complexity $\mathcal{U}_w^{(k)}$* and *lower k-Abelian complexity $\mathcal{L}_w^{(k)}$* by

$$\mathcal{U}_w^{(k)}(n) = \max_{m \leq n} \mathcal{P}_w^{(k)}(m) \qquad \text{and} \qquad \mathcal{L}_w^{(k)}(n) = \min_{m \geq n} \mathcal{P}_w^{(k)}(m).$$

These two functions can be significantly different. For example, if w is the Thue-Morse word and $k \geq 2$, then $\mathcal{U}_w^{(k)}(n) = \Theta(\log n)$ and $\mathcal{L}_w^{(k)}(n) = \Theta(1)$. This will be proved in Sect. 4.

The Abelian complexity of a binary word $w \in \{0,1\}^\omega$ can be determined using the formula

$$\mathcal{P}_w^{(1)}(n) = \max\{|u|_1 \mid u \in \mathcal{F}_n(w)\} - \min\{|u|_1 \mid u \in \mathcal{F}_n(w)\} + 1. \qquad (1)$$

For $k \in \mathbb{N}_1 \cup \{\infty\}$, let $q^{(k)} : \mathbb{N}_1 \to \mathbb{N}_1$ be the function defined by

$$q^{(k)}(n) = \begin{cases} n+1 & \text{if } n \leq 2k-1 \\ 2k & \text{if } n \geq 2k \end{cases}.$$

The significance of this function is that if w is Sturmian, then $\mathcal{P}_w^{(k)} = q^{(k)}$. This is further discussed in Sect. 3.

There are large classes of words for which the k-Abelian complexities are of the same order for many values of k. This is shown in the next two lemmas. Thus when analyzing the growth rate of the k-Abelian complexity of a word, it may be sufficient to analyze the Abelian or 2-Abelian complexity.

Lemma 3. *Let $w \in \{0,1\}^\omega$ be such that every factor of w of length k contains at most one occurrence of 1. Then $\mathcal{P}_w^{(k)}(n) = \Theta(\mathcal{P}_w^{(1)}(n))$.*

Proof. Two factors of w are k-Abelian equivalent if and only if they are Abelian equivalent and have the same prefixes and suffixes of length $k-1$. □

Lemma 4. *Let $k, m \geq 2$ and let w be a fixed point of an m-uniform morphism h. Let i be such that $m^i \geq k-1$. Then $\mathcal{P}_w^{(k)}(m^i(n+1)) = O(\mathcal{P}_w^{(2)}(n))$.*

Proof. Every factor of w of length $m^i(n+1)$ can be written as $ph^i(u)q$, where u is a factor of w of length n and $|pq| = m^i$. The k-Abelian equivalence class of $ph^i(u)q$ is determined by p, q and the 2-Abelian equivalence class of u. □

In particular, Lemma 4 can be applied to the Thue-Morse word to analyze its k-Abelian complexity once the behavior of its 2-Abelian complexity is known.

It has been shown that there are many words for which the k-Abelian and $(k+1)$-Abelian complexities are similar, but there are also many words for which they are very different. For example, there are words having bounded k-Abelian complexity but linear $(k+1)$-Abelian complexity. These words can even be assumed to be k-Abelian periodic. This is shown in the next lemma.

Lemma 5. *For every $k \geq 1$ there is a k-Abelian periodic word w such that $\mathcal{P}_w^{(k+1)}(n) = \Theta(n)$.*

Proof. Let $w \in \{0,1\}^\omega$ and let h be the morphism defined by

$$0 \mapsto 0^{k+1}10^{k-1}1, \qquad 1 \mapsto 0^k10^k1.$$

Then $h(w)$ is k-Abelian periodic and $\mathcal{P}_{h(w)}^{(k+1)}((2k+2)n) = \Theta(\mathcal{P}_w^{(1)}(n))$. The claim follows because there are words w with linear Abelian complexity. □

3 Minimal k-Abelian Complexities

In this section classes of words with small k-Abelian complexity are studied. Some well-known results about factor complexity are compared to results on k-Abelian complexity proved in [10]. It should be expected that ultimately periodic words have low complexity, and this is indeed true for k-Abelian complexity, although the k-Abelian complexity of some ultimately periodic words is higher that the k-Abelian complexity of some aperiodic words. For many complexity

measures, Sturmian words have the lowest complexity among aperiodic words. This is also true for k-Abelian complexity.

We recall the famous theorem of Morse and Hedlund [13] characterizing ultimately periodic words in terms of factor complexity. This theorem can be generalized for k-Abelian complexity: If $\mathcal{P}_w^{(k)}(n) < q^{(k)}(n)$ for some n, then w is ultimately periodic, and if w is ultimately periodic, then $\mathcal{P}_w^{(\infty)}(n)$ is bounded. This was proved in [10].

If k is finite, then this generalization does not give a characterization of ultimately periodic words, because the function $q^{(k)}$ is bounded. In fact, it is impossible to characterize ultimately periodic words in terms of k-Abelian complexity. For example, the word $0^{2k-1}1^\omega$ has the same k-Abelian complexity as every Sturmian word. On the other hand, for every ultimately periodic word w there is a finite k such that $\mathcal{P}_w^{(k)}(n) < q^{(k)}(n)$ for all sufficiently large n.

The theorem of Morse and Hedlund has a couple of immediate consequences. The words w with $\mathcal{P}_w^{(\infty)}(n) = n+1$ for all n are, by definition, Sturmian words. Thus the following classification is obtained:

- w is ultimately periodic $\Leftrightarrow \mathcal{P}_w^{(\infty)}$ is bounded.
- w is Sturmian $\Leftrightarrow \mathcal{P}_w^{(\infty)}(n) = n+1$ for all n.
- w is aperiodic and not Sturmian $\Leftrightarrow \mathcal{P}_w^{(\infty)}(n) \geq n+1$ for all n and $\mathcal{P}_w^{(\infty)}(n) > n+1$ for some n.

This can be generalized for k-Abelian complexity if the equivalences are replaced with implications:

- w is ultimately periodic $\Rightarrow \mathcal{P}_w^{(k)}$ is bounded.
- w is Sturmian $\Rightarrow \mathcal{P}_w^{(k)} = q^{(k)}$.
- w is aperiodic and not Sturmian $\Rightarrow \mathcal{P}_w^{(k)}(n) \geq q^{(k)}(n)$ for all n and $\mathcal{P}_w^{(k)}(n) > q^{(k)}(n)$ for some n.

For $k = 1$ this follows from the theorem of Coven and Hedlund [2]. For $k \geq 2$ it follows from a theorem in [10].

The above result means that one similarity between factor complexity and k-Abelian complexity is that Sturmian words have the lowest complexity among aperiodic words. Another similarity is that ultimately periodic words have bounded complexity, but the largest values can be arbitrarily high: For every n, there is a finite word u having every possible factor of length n. Then $\mathcal{P}_{u^\omega}^{(k)}(n)$ is as high as it can be for any word, i.e. the number of k-Abelian equivalence classes of words of length n.

Another direct consequence of the theorem of Morse and Hedlund is that there is a gap between constant complexity and the complexity of Sturmian words. For k-Abelian complexity there cannot be a gap between bounded complexities and $q^{(k)}$, because the function $q^{(k)}$ itself is bounded. However, the question whether there is a gap above bounded complexity is more difficult. The answer is that there is no such gap, even if only uniformly recurrent words are considered. This is proved in Sect. 5.

4 k-Abelian Complexity of the Thue-Morse Word

In this section the k-Abelian complexity of the Thue-Morse word is analyzed. Before that, the Abelian complexity of a closely related word is determined.

Let σ be the morphism defined by $\sigma(0) = 01, \sigma(1) = 00$. Let

$$S = 0100010101000100\dots$$

be the *period-doubling word*, which is the fixed point of σ, see [4].

The Abelian complexity of S is completely determined by the recurrence relations in the following lemma and by the first two values $\mathcal{P}_S^{(1)}(1) = \mathcal{P}_S^{(1)}(2) = 2$.

Lemma 6. *For $n \geq 1$,*

$$\mathcal{P}_S^{(1)}(4n - 1) = \mathcal{P}_S^{(1)}(n) + 1, \qquad\qquad \mathcal{P}_S^{(1)}(4n) = \mathcal{P}_S^{(1)}(n),$$
$$\mathcal{P}_S^{(1)}(4n + 1) = \mathcal{P}_S^{(1)}(n) + 1, \qquad\qquad \mathcal{P}_S^{(1)}(4n + 2) = \mathcal{P}_S^{(1)}(n) + 1.$$

Proof. Let $p_n = \min\{|u|_1 \mid u \in \mathcal{F}_n(S)\}$ and $q_n = \max\{|u|_1 \mid u \in \mathcal{F}_n(S)\}$. For $a \in \{0, 1\}$, $\sigma^2(a) = 010a$. Because

$$\begin{aligned}
\mathcal{F}_{4n-1}(S) = &\left\{\sigma^2(a_1 \dots a_{n-1})010 \mid a_1 \dots a_{n-1} \in \mathcal{F}_{n-1}(S)\right\} \cup \\
&\left\{10a_1\sigma^2(a_2 \dots a_n) \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{0a_1\sigma^2(a_2 \dots a_n)0 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{a_1\sigma^2(a_2 \dots a_n)01 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\},
\end{aligned}$$

$$\begin{aligned}
\mathcal{F}_{4n}(S) = &\left\{\sigma^2(a_1 \dots a_n) \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{10a_1\sigma^2(a_2 \dots a_n)0 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{0a_1\sigma^2(a_2 \dots a_n)01 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{a_1\sigma^2(a_2 \dots a_n)010 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\},
\end{aligned}$$

$$\begin{aligned}
\mathcal{F}_{4n+1}(S) = &\left\{\sigma^2(a_1 \dots a_n)0 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{10a_1\sigma^2(a_2 \dots a_n)01 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{0a_1\sigma^2(a_2 \dots a_n)010 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{a_1\sigma^2(a_2 \dots a_{n+1}) \mid a_1 \dots a_{n+1} \in \mathcal{F}_{n+1}(S)\right\},
\end{aligned}$$

$$\begin{aligned}
\mathcal{F}_{4n+2}(S) = &\left\{\sigma^2(a_1 \dots a_n)01 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{10a_1\sigma^2(a_2 \dots a_n)010 \mid a_1 \dots a_n \in \mathcal{F}_n(S)\right\} \cup \\
&\left\{0a_1\sigma^2(a_2 \dots a_{n+1}) \mid a_1 \dots a_{n+1} \in \mathcal{F}_{n+1}(S)\right\} \cup \\
&\left\{a_1\sigma^2(a_2 \dots a_{n+1})0 \mid a_1 \dots a_{n+1} \in \mathcal{F}_{n+1}(S)\right\},
\end{aligned}$$

it can be seen that

$$p_{4n-1} = p_n + n - 1, \qquad q_{4n-1} = q_n + n,$$
$$p_{4n} = p_n + n, \qquad q_{4n} = q_n + n,$$
$$p_{4n+1} = p_n + n, \qquad q_{4n+1} = q_n + n + 1,$$
$$p_{4n+2} = p_n + n, \qquad q_{4n+2} = q_n + n + 1.$$

The claim follows because $\mathcal{P}_S^{(1)}(n) = q_n - p_n + 1$ for all n. $\qquad\square$

Lemma 7. *For $n \geq 1$ and $m \geq 0$,*

$$\mathcal{P}_S^{(1)}(n) = O(\log n), \quad \mathcal{P}_S^{(1)}((2 \cdot 4^m + 1)/3) = m + 2, \quad \mathcal{P}_S^{(1)}(2^m) = 2.$$

Proof. Follows from Lemma 6 by induction. $\qquad\square$

The Abelian complexity of S has a logarithmic upper bound and a constant lower bound. These bounds are the best possible increasing bounds.

Theorem 8. $\mathcal{U}_S^{(1)}(m) = \Theta(\log n)$ *and* $\mathcal{L}_S^{(1)}(m) = 2$.

Proof. Follows from Lemma 7. $\qquad\square$

Now, let τ be the Thue-Morse morphism defined by $\tau(0) = 01, \tau(1) = 10$. Let

$$T = 0110100110010110\ldots$$

be the Thue-Morse word, which is a fixed point of τ. The first few values of $\mathcal{P}_T^{(2)}$ are

$$2, 4, 6, 8, 6, 8, 10, 8, 6, 8, 8, 10, 10, 10, 8, 8, 6, 8, 10, 10.$$

The 2-Abelian equivalence of factors of T can be determined with the help of the following lemma.

Lemma 9. *Words $u, v \in \{0, 1\}^*$ are 2-Abelian equivalent if and only if*

$$|u| = |v|, \qquad |u|_{00} = |v|_{00}, \qquad |u|_{11} = |v|_{11} \qquad and \qquad \mathrm{pref}_1(u) = \mathrm{pref}_1(v).$$

Proof. The "only if" direction follows immediately from the definition of 2-Abelian equivalence. For the other direction, it follows from the assumptions that $|u|_{01} + |u|_{10} = |v|_{01} + |v|_{10}$. In any word $w \in \{0, 1\}^*$, $|w|_{01}$ and $|w|_{10}$ can differ by at most one. If $|w|_{01} + |w|_{10}$ is even, then $|w|_{01} = |w|_{10}$. If it is odd and $\mathrm{pref}_1(w) = 0$, then $|w|_{01} = |w|_{10} + 1$. If it is odd and $\mathrm{pref}_1(w) = 1$, then $|w|_{01} + 1 = |w|_{10}$. This means that $|u|_{01} = |v|_{01}$ and $|u|_{10} = |v|_{10}$ and u and v are 2-Abelian equivalent. $\qquad\square$

The following lemma states that if u is a factor of T, then the numbers $|u|_{00}$ and $|u|_{11}$ can differ by at most one.

Lemma 10. *In the image of any word under τ, between any two occurrences of 00 there is an occurrence of 11 and vice versa.*

Proof. 00 can only occur in the middle of $\tau(10)$ and 11 can only occur in the middle of $\tau(01)$. The claim follows because 10's and 01's alternate in all binary words. □

Let u be a factor of T. If $|u|$ and $|u|_{00} + |u|_{11}$ are given, then there are at most 4 possibilities for the 2-Abelian equivalence class of u. This is stated in a different way in the next lemma. First we define a function ϕ as follows. If $w = a_1 \ldots a_n$, then $\phi(w) = b_1 \ldots b_{n-1}$, where $b_i = 0$ if $a_i a_{i+1} \in \{01, 10\}$ and $b_i = 1$ if $a_i a_{i+1} \in \{00, 11\}$. If $w = a_1 a_2 \ldots$ is an infinite word, then $\phi(w) = b_1 b_2 \ldots$ is defined in an analogous way.

Lemma 11. *Let $u_1, \ldots, u_n$ be factors of T. If $\phi(u_1), \ldots, \phi(u_n)$ are Abelian equivalent, then $u_1, \ldots, u_n$ are in at most 4 different 2-Abelian equivalence classes.*

Proof. The numbers $|u_i|_{00} + |u_i|_{11} = |\phi(u_i)|_1$ are equal for all i; let this number be m. By Lemma 10, $\{|u_i|_{00}, |u_i|_{11}\} = \{\lfloor m/2 \rfloor, \lceil m/2 \rceil\}$. There are at most four different possible values for the triples $(|u_i|_{00}, |u_i|_{11}, \mathrm{pref}_1(u_i))$. The claim follows from Lemma 9. □

Now it can be proved that the 2-Abelian complexity of T is of the same order as the Abelian complexity of $\phi(T)$. It is known that $\phi(T)$ is actually the period-doubling word S.

Lemma 12. *For $n \geq 2$, $\mathcal{P}_S^{(1)}(n-1) \leq \mathcal{P}_T^{(2)}(n) \leq 4\mathcal{P}_S^{(1)}(n-1)$.*

Proof. If the factors of T of length n are $u_1, \ldots, u_m$, then the factors of $\phi(T)$ of length $n-1$ are $\phi(u_1), \ldots, \phi(u_m)$. If u_i and u_j are 2-Abelian equivalent, then $\phi(u_i)$ and $\phi(u_j)$ are Abelian equivalent, so the first inequality follows. The second inequality follows from Lemma 11 □

Lemma 13. *For $n \geq 1$ and $m \geq 0$,*

$$\mathcal{P}_T^{(2)}(n) = O(\log n), \quad \mathcal{P}_T^{(2)}((2 \cdot 4^m + 4)/3) = \Theta(m), \quad \mathcal{P}_T^{(2)}(2^m + 1) \leq 8.$$

Proof. Follows from Lemmas 12 and 8. □

The k-Abelian complexity of T behaves in a similar way as the Abelian complexity of S.

Theorem 14. *Let $k \geq 2$. Then*

$$\mathcal{U}_T^{(k)}(m) = \Theta(\log n) \qquad and \qquad \mathcal{L}_T^{(k)}(m) = \Theta(1).$$

Proof. Follows from Lemmas 13 and 4. □

5 Arbitrarily Slowly Growing k-Abelian Complexities

In this section it is studied whether there is a gap above bounded k-Abelian complexity. This question can be formalized in two ways:

- Does there exists an increasing unbounded function $f : \mathbb{N}_1 \to \mathbb{N}_1$ such that for every infinite word w either $\mathcal{P}_w^{(k)}$ is bounded or $\mathcal{P}_w^{(k)} = \Omega(f)$?
- Does there exists an increasing unbounded function $f : \mathbb{N}_1 \to \mathbb{N}_1$ such that for every infinite word w either $\mathcal{P}_w^{(k)}$ is bounded or $\mathcal{P}_w^{(k)} \neq O(f)$?

The first question has already been answered negatively in Sect. 4. The answer to the second question is also negative, even if only uniformly recurrent words are considered. A construction proving this is given below.

Let $n_1, n_2, \ldots$ be a sequence of integers greater than 1. Let $m_j = n_1 \ldots n_j$ for $j = 0, 1, 2, \ldots$. Let $a_i = 0$ if the greatest j such that $m_j | i$ is even and $a_i = 1$ otherwise. Let $U = a_1 a_2 a_3 \ldots$.

Lemma 15. *The word U is uniformly recurrent.*

Proof. For every factor u of U, there is a j such that u is a factor of $\mathrm{pref}_{m_j-1}(U)$. Because $U \in \{\mathrm{pref}_{m_j-1}(U)0, \mathrm{pref}_{m_j-1}(U)1\}^\omega$, every factor of U of length $m_j + |u| - 1$ contains u. $\qquad\square$

Lemma 16. *For every n, let n' be such that $m_{n'-1} < n \leq m_{n'}$. Then*

$$\mathcal{P}_U^{(1)}(n) \leq n' + 1.$$

For all $J \geq 1$, if $n = 2\sum_{j=1}^{J}(m_{2j} - m_{2j-1})$, then

$$\mathcal{P}_U^{(1)}(n) \geq \frac{n'+1}{2}.$$

For all $j \geq 1$,

$$\mathcal{P}_U^{(1)}(m_j) = 2.$$

Proof. Formula (1) will be used repeatedly in this proof. Another important simple fact is that if a, b, c are integers and c divides a, then $\lfloor (a+b)/c \rfloor = a/c + \lfloor b/c \rfloor$.

For all $n \geq 1$,

$$|\mathrm{pref}_n(U)|_1 = \sum_{i=1}^{\infty} (-1)^{i+1} \left\lfloor \frac{n}{m_i} \right\rfloor,$$

and for all $n \geq 1$ and $l \geq 0$,

$$|\mathrm{rfact}_n^l(U)|_1 = |\mathrm{pref}_{n+l}(U)|_1 - |\mathrm{pref}_l(U)|_1 = \sum_{i=1}^{\infty} (-1)^{i+1} \left(\left\lfloor \frac{n+l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor \right).$$

For all i,

$$\left\lfloor \frac{n+l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor \in \left\{ \left\lfloor \frac{n}{m_i} \right\rfloor, \left\lceil \frac{n}{m_i} \right\rceil \right\}.$$

Moreover, for every l there is an i' such that

$$\left\lfloor \frac{n+l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor = \begin{cases} 1 & \text{if } n' \le i \le i' \\ 0 & \text{if } i \ge i' \end{cases},$$

so

$$\sum_{i=n'}^{\infty} (-1)^{i+1} \left(\left\lfloor \frac{n+l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor \right) \in \left\{ 0, (-1)^{n'+1} \right\}.$$

Thus there are at most $n'+1$ possible values for $|\mathrm{rfact}_n^l(U)|_1$ and $\mathcal{P}_U^{(1)}(n) \le n'+1$.

Consider the second claim. Let $n = 2\sum_{j=1}^{J}(m_{2j} - m_{2j-1})$ and $l = m_{2J+1} - n/2$. Then

$$|\mathrm{rfact}_n^l(U)|_1 - |\mathrm{pref}_n(U)|_1 = \sum_{i=1}^{\infty} (-1)^{i+1} \left(\left\lfloor \frac{n+l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor - \left\lfloor \frac{n}{m_i} \right\rfloor \right)$$

and for $i \le 2J$

$$\left\lfloor \frac{n+l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor - \left\lfloor \frac{n}{m_i} \right\rfloor$$

$$= \frac{m_{2J+1} + \sum_{(i+1)/2 \le j \le J}(m_{2j} - m_{2j-1})}{m_i} + \left\lfloor \frac{\sum_{1 \le j < (i+1)/2}(m_{2j} - m_{2j-1})}{m_i} \right\rfloor$$

$$- \frac{m_{2J+1} - \sum_{(i+1)/2 \le j \le J}(m_{2j} - m_{2j-1})}{m_i} - \left\lfloor -\frac{\sum_{1 \le j < (i+1)/2}(m_{2j} - m_{2j-1})}{m_i} \right\rfloor$$

$$- \frac{2\sum_{(i+1)/2 \le j \le J}(m_{2j} - m_{2j-1})}{m_i} - \left\lfloor \frac{2\sum_{1 \le j < (i+1)/2}(m_{2j} - m_{2j-1})}{m_i} \right\rfloor$$

$$= \left\lfloor \frac{s}{m_i} \right\rfloor - \left\lfloor -\frac{s}{m_i} \right\rfloor - \left\lfloor \frac{2s}{m_i} \right\rfloor,$$

where $s = \sum_{1 \le j < (i+1)/2}(m_{2j} - m_{2j-1})$. If i is even, then $m_i/2 \le s < m_i$, and if i is odd and $i > 1$, then $m_{i-1}/2 \le s < m_{i-1}$. Thus

$$\left\lfloor \frac{s}{m_i} \right\rfloor - \left\lfloor -\frac{s}{m_i} \right\rfloor - \left\lfloor \frac{2s}{m_i} \right\rfloor = \begin{cases} 0 & \text{if } i \text{ is even or } i = 1 \\ 1 & \text{if } i \text{ is odd and } i > 1 \end{cases}$$

and

$$\mathcal{P}_U^{(1)}(n) \ge |\mathrm{rfact}_n^l(U)|_1 - |\mathrm{pref}_n(U)|_1 + 1$$

$$= \sum_{i'=2}^{J} (-1)^{(2i'-1)+1} + \sum_{i=2J+1}^{\infty} (-1)^{i+1} \left(\left\lfloor \frac{n+l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor - \left\lfloor \frac{n}{m_i} \right\rfloor \right) + 1$$

$$= J + 1 = \frac{n'+1}{2}.$$

Consider the third claim. Let $n = m_j$. Then

$$|\text{rfact}_n^l(U)|_1 = \sum_{i=1}^{\infty} (-1)^{i+1} \left(\left\lfloor \frac{m_j + l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor \right)$$

$$= \sum_{i=1}^{j} (-1)^{i+1} \frac{m_j}{m_i} + \sum_{i=j+1}^{\infty} (-1)^{i+1} \left(\left\lfloor \frac{m_j + l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor \right)$$

and

$$\sum_{i=j+1}^{\infty} (-1)^{i+1} \left(\left\lfloor \frac{m_j + l}{m_i} \right\rfloor - \left\lfloor \frac{l}{m_i} \right\rfloor \right) \in \{0, (-1)^j\},$$

so $\mathcal{P}_U^{(1)}(n) = 2$. $\qquad\square$

If $n_i = 2$ for all i, then the word U is the period-doubling word S. Thus Lemma 16 gives an alternative proof for Theorem 8.

Theorem 17. *For every increasing unbounded function $f : \mathbb{N}_1 \to \mathbb{N}_1$ there is a uniformly recurrent word $w \in \{0,1\}^\omega$ such that $\mathcal{P}_w^{(k)}(n) = O(f(n))$ but $\mathcal{P}_w^{(k)}(n)$ is not bounded.*

Proof. Follows from Lemmas 3, 15 and 16. $\qquad\square$

6 Conclusion

In this paper we have investigated some generalizations of the results of Morse and Hedlund and those of Coven and Hedlund for k-Abelian complexity. We have pointed out many similarities but also many differences. We have studied the k-Abelian complexity of the Thue-Morse word and proved that there are uniformly recurrent words with arbitrarily slowly growing k-Abelian complexities.

There are many open question and possible directions for future work. One open problem related to Lemma 5 is to determine the maximal $(k+1)$-Abelian complexity of a k-Abelian periodic word. Another interesting topic would be the k-Abelian complexities of morphic words. For example, Theorem 17 does not hold if morphic words are considered instead of uniformly recurrent words, because the k-Abelian complexity of a morphic word is always a computable function. But for a morphic (or pure morphic) word w, how slowly can $\mathcal{U}_w^{(k)}(n)$ grow without being bounded? Can it grow slower than logarithmically? More generally, can the possible k-Abelian complexities of some subclass of morphic words be classified?

References

1. Julien Cassaigne, Gwenael Richomme, Kalle Saari, and Luca Q. Zamboni. Avoiding Abelian powers in binary words with bounded Abelian complexity. *Internat. J. Found. Comput. Sci.*, 22(4):905–920, 2011.

2. Ethan M. Coven and G. A. Hedlund. Sequences with minimal block growth. *Math. Systems Theory*, 7:138–153, 1973.

3. James Currie and Narad Rampersad. Recurrent words with constant Abelian complexity. *Adv. in Appl. Math.*, 47(1):116–124, 2011.

4. David Damanik. Local symmetries in the period-doubling sequence. *Discrete Appl. Math.*, 100(1–2):115–121, 2000.

5. F. M. Dekking. Strongly nonrepetitive sequences and progression-free sets. *J. Combin. Theory Ser. A*, 27(2):181–185, 1979.

6. Mari Huova and Juhani Karhumäki. Observations and problems on k-abelian avoidability. In *Combinatorial and Algorithmic Aspects of Sequence Processing (Dagstuhl Seminar 11081)*, pages 2215–2219, 2011.

7. Mari Huova, Juhani Karhumäki, and Aleksi Saarela. Problems in between words and abelian words: k-abelian avoidability. *Theoret. Comput. Sci.*, 454:172–177, 2012.

8. Mari Huova, Juhani Karhumäki, Aleksi Saarela, and Kalle Saari. Local squares, periodicity and finite automata. In Cristian Calude, Grzegorz Rozenberg, and Arto Salomaa, editors, *Rainbow of Computer Science*, pages 90–101. Springer, 2011.

9. Juhani Karhumäki, Svetlana Puzynina, and Aleksi Saarela. Fine and Wilf's theorem for k-abelian periods. In *Proceedings of the 16th International Conference on Developments in Language Theory*, pages 296–307, 2012.

10. Juhani Karhumäki, Aleksi Saarela, and Luca Zamboni. On a generalization of Abelian equivalence and complexity of infinite words, Submitted.

11. Veikko Keränen. Abelian squares are avoidable on 4 letters. In *Proceedings of the 19th International Colloquium on Automata, Languages and Programming*, pages 41–52, 1992.

12. Robert Mercaş and Aleksi Saarela. 3-abelian cubes are avoidable on binary alphabets, Submitted.

13. Marston Morse and Gustav A. Hedlund. Symbolic dynamics. *Amer. J. Math.*, 60(4):815–866, 1938.

14. Marston Morse and Gustav A. Hedlund. Symbolic dynamics II: Sturmian trajectories. *Amer. J. Math.*, 62(1):1–42, 1940.

15. Svetlana Puzynina and Luca Q. Zamboni. Abelian returns in Sturmian words. *J. Combin. Theory, Ser. A*, 120(2):390–408, 2013.

16. Gwenael Richomme, Kalle Saari, and Luca Q. Zamboni. Balance and Abelian complexity of the Tribonacci word. *Adv. in Appl. Math.*, 45(2):212–231, 2010.

17. Gwenael Richomme, Kalle Saari, and Luca Q. Zamboni. Abelian complexity of minimal subshifts. *J. Lond. Math. Soc. (2)*, 83(1):79–95, 2011.

18. Aleksi Saarela. Ultimately constant abelian complexity of infinite words. *J. Autom. Lang. Comb.*, 14(3–4):255–258, 2009.